

The Ecologic Method in the Study of Environmental Health. I. Overview of the Method

by Stephen D. Walter*

This paper summarizes the salient features of the ecologic method, with emphasis on its application in the study of environmental health. Various types of ecologic design are described, with examples. Finally, the main advantages and disadvantages are indicated. A companion paper discusses the methodology of ecologic designs in more detail and describes a census of data sets with potential suitability for the ecologic study of water quality and human health.

Introduction

This paper gives an overview of various ecologic study designs, with emphasis on studies of environmental effects on human health. The next section gives a general description of ecologic studies, in the context of epidemiology. The third section describes some of the main types of ecologic design. The fourth section lists some advantages and disadvantages of the ecologic approach, in comparison to other epidemiologic options. A companion paper develops the methodologic issues more fully and presents a census of U.S. and Canadian data sets with potential applicability in the study of water quality and human health.

General Description of Ecologic Studies

The unique distinguishing feature of an ecologic study using epidemiologic data is that its unit of analysis is a group of individuals. This is in contrast to all other epidemiologic designs, where information is available at the level of the individual person in the population. The loss of information through ecologic aggregation is important because special care is required for the interpretation of ecologic associations with postulated risk factors. Some of the potential biases affecting ecologic studies are described below. Despite these biases, there are a number of advantages of ecologic studies over other epidemiologic designs; these include the ability to study large populations at relatively low cost and to address questions of environmental health that might be difficult or impossible to study with other approaches.

A generic example of an ecologic situation might arise as follows. Suppose we are interested in the possible association of a water contaminant (which will be denoted by X) and a health outcome (to be denoted by Y). If it were feasible to do so, such an association might be investigated epidemiologically using a cohort design. With the cohort method, individual members of

a population are enrolled into the study, and their exposure to X is ascertained at baseline and monitored over a period of time. Similarly, disease events Y occurring in the population are also ascertained prospectively over time. By assembling suitable subgroups of individuals with similar levels of exposure to X , one can estimate and compare their risks of Y in a certain period of time. Important to note is that we have taken individual exposure levels into account.

In contrast, an ecologic approach to the same problem would not have individual linkage of information on X and Y . Instead, we might choose to study the problem by identifying the level of exposure to X in the water supplies of various communities within the population. We would also estimate the rate of health events Y in the same communities. The analysis of ecologic data of this type is then intended to assess the association between X and Y on a community basis rather than at the individual level.

Many of the same concerns of methodologic quality and validity of data apply to both the cohort design and the ecologic design for this type of problem. For instance, we want to be assured that the laboratory method for the measurement of X in water samples is accurate. Also, we would require that all health events in the population are identified and recorded in a consistent and unbiased manner. Finally, we would need to consider the possibility of other exposure variables (related to water quality or otherwise) that might have a confounding effect on the apparent association between X and Y .

However, the key methodologic difference between the association as measured in cohort or ecologic data is that the ecologic design provides no information at all on the joint distribution of X and Y at the individual level. In particular, there is no assurance that individuals experiencing the health event Y were indeed those who were exposed to X . In an ecologic study, persons are assigned to community subgroups of the population on the basis of residence information, derived, for instance, from municipal tax assessment rolls. However, they may spend part or all of their time elsewhere, and so may consume water with a

*Department of Clinical Epidemiology and Biostatistics, McMaster University Medical Centre, Hamilton, Ontario, Canada, L8N 3Z5.

different level of X . Even within communities, there may be variation in the level of individual exposure to X , either because of differential mixing in the water supply system, or because of the use of alternative supplies such as bottled spring water or devices such as water softeners. In an ecologic study, one has no alternative but to assume that the same level of exposure X applies to all members of the ecologic population subgroup. The extent to which this is a valid assumption will depend on the size of the population subgroups and their heterogeneity, and possibly other factors, as described below.

The two main uses of the ecologic design in epidemiology are the generation/testing of etiologic hypotheses and the evaluation of health interventions. Typical examples of etiologic investigations include assessments of environmental contaminants and their relationship to health outcomes, or the relationship of "natural" exposure to health, for instance the association of water hardness with cardiovascular mortality (1) or the association of asbestos cement water piping and cancer (2). Examples of ecologic intervention studies include the MRFIT study (3) to evaluate health education and intervention on risk factors for coronary heart disease and the relationship of cervical pap smear screening to reduction in cervical cancer mortality (4). This paper will concentrate primarily on the etiologic type of investigation; this is the most active area of research concerning environmental correlates of health. Examples of intervention studies in this area are rather few in number; one example is the investigations of the health effects of fluoridation of the water supply (5,6). We begin by classifying ecologic designs, drawing heavily on the work of Morgenstern (7).

Types of Ecologic Design

Exploratory Ecologic Studies

An exploratory ecologic study usually examines the spatial variation in disease rates, but without any direct incorporation of exposure information. Typical examples are investigations based on cancer atlases; here the rates for cancers of interest would be examined for evidence of spatial autocorrelation, i.e., the tendency for rates to be clustered geographically. Such clustering tendencies might be related to environmental exposure variables, such as water or air quality. The analyses may be informal "eyeball" assessments of the maps, or could involve formal statistical tests for spatial autocorrelation, such as the rank adjacency method (8) or the Moran coefficient (9). Because exposure information is not directly incorporated into the analysis, this type of study is usually hypothesis generating rather than hypothesis testing.

An example of this kind of study is that by Savitz and Redmond (10), who studied the incidence of cancer in Pennsylvania. They defined 30 geographic areas of between 6 and 39 census tracts each and evaluated the fit of the data to a product model involving age and area effects. The objective of the analysis was to identify discrepancies between the data and the model predictions, which might indicate different age-specific effects within certain geographic areas.

Multigroup Comparison Ecologic Study

In the multigroup comparison design, data on exposure to X and the health outcome Y are collected on a group basis for several regions. For instance, one might measure the hardness of the water supply in a number of communities and the corresponding mortality rates from ischemic heart disease. The objective of the statistical analysis is then to decide if any association between X and Y is statistically significant and substantively meaningful, allowing for possible bias or confounding. The preferred analysis for this kind of data is regression rather than correlation (7). Regression allows the estimation of the relative risk associated with changes in exposure to X ; under ideal circumstances this relative risk will be the same as that which would have been estimated in individually linked data.

An example of an ecologic study where the geographic subgroups were census tracts is that of Carlo and Mettlin (11), investigating site-specific cancer rates in Erie County, New York. An example of a study using municipality as the unit of analysis is Isacson's (12) study in Iowa.

Time-Trend Ecologic Studies

In time-trend ecologic studies, a single population is assessed with respect to its changes over time in the rates of a disease Y and the corresponding changes in exposure X over the same period of time. An association between X and Y would be suggested if changes in X are paralleled by similar changes in Y .

In practice, it is often difficult to find populations that have experienced substantial changes in X over time, other than situations where a gradual increase or a gradual decrease has occurred. If the change in exposure has been uniformly monotonic, then it may be more difficult to identify the corresponding point in time where Y has changed.

An example of this problem is the relationship between the death rate from respiratory tuberculosis (TB) and the introduction of chemotherapy. As indicated by McKeown (13), the death rate from TB has been steadily declining since the 1830s; the tubercle bacillus was identified in the 1880s and chemotherapy was introduced in the 1940s. The incremental decrease in TB death rates that might be associated with the introduction of chemotherapy is thus hard to identify.

Another example of this kind is the change in the mortality rate from cervical cancer following the introduction of pap smear screening. Time series data from Scandinavia indicate how the rate had changed following the introduction of screening to various parts of the population (4). Because the rate of cervical cancer was declining before the introduction of screening, it is once again more difficult to clearly delineate the effect of screening.

A further difficulty in the time series approach is that it may be necessary to allow for latency in the exposure. For instance, the effect of many carcinogens is not felt for many years following exposure. Many occupational cancers are associated with workplace hazards that were experienced 20 to 30 years earlier, and the situation is likely to be similar for environmental hazards. One would have to correlate changes in the health outcome Y with changes in the environmental exposure X (e.g., a water quality variable) that had occurred sometime previously. A problem is that one usually has no precise estimate of what the appropriate latent period might be. Also, if one assumed that the latent period

was, say, 20 years, one would have to ignore health information on *Y* for which the corresponding exposure information 20 years earlier was not available; the net effect will be to shorten (possibly by a considerable amount) the useable length of the time trend data on *Y*. A further practical difficulty if latency applies is that it is correspondingly more difficult to identify the appropriate population members who were exposed to *X* in previous years.

Multiple Group Time-Trend Ecologic Studies

The multiple group time-trend study is a mixture of the multigroup comparison study and single group time-trend study. In it one identifies changes over time in both the exposure rate and the disease outcome rate for several population subgroups. An example is the study by Crawford et al. (14), who investigated the changes in water hardness in several communities and the corresponding changes in the rate of coronary heart disease.

In general, the multigroup time-trend design is stronger than the single group time-trend design because its results are less susceptible to confounding. It is relatively unlikely that the same confounding variable could lead to a spurious ecologic association in a set of time series, relative to the chance of this happening in a single time series. The use of multiple time series is a form of replication that brings greater plausibility to the scientific results.

Advantages and Disadvantages of Ecologic Studies

Advantages

The main advantage of the ecologic approach is that it allows the study of very large populations. Because exposure and health information are used on a group basis, there is a considerable increase in cost efficiency as compared to designs where individual data are required. Alternative designs such as the case-control method typically involve samples of at most several hundred cases and controls; the typical prospective cohort design might involve at most several thousand individuals. But an ecologic design is capable of studying populations that are orders of magnitude larger. Ecologic studies have even been done to make international comparisons, thereby including populations of many millions. An example is an analysis of the relationship of coronary heart disease mortality to the polyunsaturated/saturated fat ratio in the diet of approximately 20 countries (15).

Another practical advantage of many ecologic studies is that they use existing databases. For instance, if water quality data are routinely available in a particular geographic area, and if disease outcomes (e.g., incident cases of cancer) are recorded in a registry, then the two sources of data may be used directly, without the necessity for contact with individual population members.

Both the ability to study large populations and the frequent use of available data imply that the ecologic design may be one of the most cost-efficient epidemiologic approaches. Further cost savings may result because it is often possible to execute an ecologic study in a relatively short period of time. There is no necessity to await the occurrence of incident cases of disease, as is required in a cohort study; similarly, there is no need to wait for a case series of sufficient magnitude to accrue, as is required in case-control studies.

Because large populations can be studied using ecologic designs, one may investigate relatively small increases in risk. Environmental exposures that are associated with small or moderate increases in risk, but which apply to large segments of the population, are capable of generating quite large numbers of cases of disease. Such factors can be of great significance to public health. The overall impact of such exposures can be conveyed numerically by use of the population attributable risk index, which represents the proportion of all cases of disease in a population that might be associated with exposure (16,17). The population attributable risk is a function both of the relative risk of individuals exposed versus not exposed to the hazard in question and of the proportion of the population which is exposed. It is possible for the population attributable risk to attain quite high values when the exposure prevalence rate is high, even though the relative risk is only modest (18). However, in order to demonstrate the statistical significance of a small relative risk, large populations must be studied. The ecologic design is often well suited for this purpose.

An example of the "small risk, large population" scenario is that of low-level carcinogenicity in well water. Crump and Guess (19) have calculated an upper limit on the risk for all carcinogens identified in well water in the United States. This is an estimated 0.1% increase in lifetime excess risk for all cancers, and less than a 10% increase in the number of cases of rectal, colon, or bladder cancer individually. Crump and Guess concluded that epidemiologic studies may overestimate the effect of drinking water on cancer rates, possibly because of confounding with other environmental risk factors not measured, because of collinearity between organic concentrations in water and other factors in the environment, or because humans are more susceptible than animal species tested for carcinogenicity of the same contaminants. They have concluded that "increased risks of rectal, bladder, and colon cancer of the magnitude suggested by these studies are large enough to be of concern yet small enough to be very difficult to separate from confounding risks associated with other environmental risk factors" (19).

Another advantage of the ecologic approach is its usefulness in the investigation of suspicious clusters of disease in relatively small geographic areas. Examples of this type include studies of apparent increases in cancer rates near locally contaminated water supplies. Communities that suspect they are experiencing sudden or sustained increases in health event rates often demand that epidemiologic investigations be carried out. Examples of this kind include an investigation of an outbreak of leukemia associated with industrially contaminated ground water in Woburn, Massachusetts (20), and the Upper Ottawa Street Landfill Study in Hamilton, which investigated the health of residents near a landfill site, possibly subjected to airborne and waterborne contaminants (21).

A common feature of investigations of local health problems is that a suitable comparison must be made to an appropriate control group of individuals not exposed to the hazard in question. Some of these studies involve the use of mortality or cancer incidence registry data and can therefore be completely ecologic in nature, without requiring contact with the individuals in the study area. However, in practice the ecologic information is often supplemented with personal interviews concerning health and/or exposure to the postulated contaminant. If questionnaire or

other individual data are used, the study ceases to be one with a pure ecologic design but assumes mixed design.

Disadvantages

The strongest disadvantage of the ecologic design arises because of its inherent feature of using aggregated data. Because the joint distribution of exposure and health at the individual level remains unknown, there is the possibility that the so-called "ecologic fallacy" would apply; this fallacy is described in the companion paper in more detail, but in summary we may say that it leads to possible distortion of the association between exposure and disease. It is possible for variables *X* and *Y* to be apparently associated in ecologic data, when no association exists at the individual level; similarly, it is possible that two variables *X* and *Y* which are correlated at the individual level show no association when studied in aggregated data. By careful attention to methodologic issues in the design of ecologic studies, it may be possible to minimize the effects of the ecologic fallacy; however, it is usually difficult to assess the likelihood of an ecologic fallacy having occurred once a study has been completed.

The possibility of fallacious ecologic associations has led many epidemiologists to be critical of the ecologic method. Most would agree that it is generally preferable to use a nonecologic design if this is feasible. At the same time, if an ecologic design is selected, it requires considerable attention to methodologic rigor in order to minimize the potential ecologic fallacy problem.

A second disadvantage of the ecologic approach is more practical in nature. If existing databases are to be used with the ecological method, then obviously one is limited by the extent of those databases. The use of routinely collected laboratory data on water quality will by necessity restrict attention to those variables that have been measured. These variables may or may not include the most relevant quantities for health investigations; specific carcinogens or bacteria may not have been explicitly measured and so cannot be studied.

The same type of limitation may apply to routinely available health data. Disease registries may not include disease events of interest or may classify them with coding schemes that are inappropriate to the research study question. Mortality data, for instance, are ascertained on almost 100% of deaths, but the coded cause of death may not always be accurate. In addition, one might be interested in contributory causes of death, rather than underlying causes of death, and these may be difficult to extract from routine vital statistics.

For less serious health events such as nonfatal gastrointestinal disorders, there may be no suitable disease registry or database available at all. Much of this type of morbidity may go completely unrecorded if individuals do not seek health care. Even if they do seek care, the information they provide may be widely dispersed in physicians' notes or hospital admission forms, and therefore difficult to access for a large population group. Generally speaking, it is more likely that serious health events (such as diagnosis of cancer or death) will be recorded in a centralized database, whereas data on minor morbidity will either not be recorded at all or recorded in a nonsystematic way on a noncentralized basis.

There may be further difficulty in drawing causal conclusions from ecologic data because of possible confounding. For example, if an association is found of increased health events with poorer water quality, then it is possible that the association is due

to confounding with socioeconomic status. If persons of low socioeconomic status tend to reside in regions where public services in general and water quality in particular are poorer, then an apparent association of health with poor water quality would be induced through the general effects of a social class gradient for several diseases. If this (possibly hypothetical) scenario applied, the true risk factor would be low socioeconomic status rather than poor water quality. It would be difficult for an ecologic analysis to separate the effect of water quality and low socioeconomic status and might falsely conclude that water quality was indeed the causal variable.

Conclusion

This paper has described several types of ecologic study designs, with examples. The following companion paper discusses the various methodologic issues involved in more depth. It also considers the practical applicability of the method by using a census of U.S. and Canadian data sets on water quality and human health.

An earlier version of this paper was written under contract to the Committee on the Assessment of the Human Health Effects of Great Lakes Water Quality, International Joint Commission.

REFERENCES

1. Shaper, A. G., Packham, R. F., and Pocock, S. J. The British regional heart study: cardiovascular mortality and water quality. *J. Environ. Pathol. Toxicol.* 4(2): 89-111 (1980).
2. Meigs, J. W., Walter, S. D., and Heston, J. F. Asbestos cement pipe and cancer in Connecticut 1955-1974. *J. Environ. Health* 42: 187-191 (1980).
3. Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial; risk factor changes and mortality results. *J. Am. Med. Assoc.* 248: 1465-1477 (1982).
4. Hakama, M. Trends in the incidence of cervical cancer in the Nordic Countries. In: *Trends in Cancer Incidence* (K. Magnus, Ed.), Hemisphere Publishing, Washington, DC, 1982, pp. 279-292.
5. Chilvers, C. Cancer mortality by site and fluoridation of water supplies. *J. Epidemiol. Commun. Health* 36: 237-242 (1982).
6. Chilvers, C. and Deirdre, C. Cancer mortality in England in relation to levels of natural occurring fluoride in water supplies. *J. Epidemiol. Commun. Health* 39: 44-47 (1985).
7. Morgenstern, H. Uses of ecologic analysis in epidemiologic research. *Am. J. Public Health* 72: 1336-1343 (1982).
8. Smans, M. Analysis of spatial aggregation. In: *Cancer Mapping* (P. Boyle, C. S. Muir, and E. Grundmann, Eds.), Springer-Verlag, Paris, 1989, pp. 253-268.
9. Cliff, A. D., and Ord, J. K. *Spatial Process Models and Applications*. Pion Limited, London, 1981.
10. Savitz, S. A., and Redmond, C. K. Screening for geographic heterogeneity of disease rates: application to cancer incidence in Allegheny County, Pennsylvania, 1969-1971. *J. Chron. Dis.* 38: 145-156 (1985).
11. Carlo, G. L., and Mettlin, C. J. Cancer incidence and trichloromethane concentrations in a public drinking water system. *Am. J. Public Health* 70: 523-525 (1980).
12. Isacson, P., Bean, J. A., Splinter, R., Olson, K. B., and Kohler, J. Drinking water and cancer incidence in Iowa. *Am. J. Epidemiol.* 121: 856-869 (1985).
13. McKeown, T. *The Role of Medicine: Dream, Mirage or Nemesis*. Nuffield Provincial Hospitals Trust, London, 1976.
14. Crawford, M. D. Changes in water hardness and local death rates. *Lancet* ii: 327-329 (1971).
15. Rose, G. Causes of the trends and variations in CHD mortality in different countries. *Int. J. Epidemiol.* 18: S174-S179 (1989).
16. Walter, S. D. The estimation and interpretation of attributable risk in health research. *Biometrics* 32: 829-849 (1976).

17. Walter, S. D. Calculation of attributable risks from epidemiological data. *Int. J. Epidemiol.* 7: 175–182 (1978).
18. Lilienfeld, A. M. *Foundation of Epidemiology*. Oxford University Press, New York, 1976.
19. Crump, K. S., and Guess, H. A. Drinking water and cancer: review of recent epidemiological findings and assessment of risks. *Annu. Rev. Public Health* 3: 339–357 (1982).
20. Lagakos, S. W., Wesson, B. J., and Zelen, M. An analysis of contaminated well water and health effects in Woburn, Massachusetts. *J. Am. Stat. Assoc.* 81: 583–614 (1986).
21. Hertzman, C., Hayes, M., Singer, J., and Highland, J. Upper Ottawa street landfill site health study. *Environ. Health Perspect.* 75: 173–195 (1987).